УДК 81`33

# О НЕКОТОРЫХ ЛИНГВИСТИЧЕСКИХ АСПЕКТАХ РАСПОЗНАВАНИЯ РЕЧИ

**Тачеев П.С.**
**научный руководитель старший преподаватель Николаева Н.В.**
***Сибирский федеральный университет***

Человеческая речь – основное средство общения между людьми. Она способствует быстрому и однозначному обмену информацией, что в свою очередь приводит к значительному накоплению знаний и к такому же стремительному развитию человечества. В современном компьютеризированном мире, количество информации намного превышает то, с которым ранее сталкивался человек. В таких условиях появилась потребность в быстром и качественном способе ввода информации. Попытки «научить» компьютер понимать и симулировать человеческую речь появились на заре эры ЭВМ. Сегодня над этой задачей работают уже десятки крупнейших компаний и корпораций для того, чтобы изменить наше представление о способах взаимодействия с компьютером.

Как правило большинство систем распознавания речи делятся на два типа, это - те, которые распознают по заранее записанному образцу, и те, которые выделяют из общего потока речи отдельные лексические элементы. Также можно разделить все системы на два класса: на зависимые от диктора (системы голосового управления) и независимые (системы диктовки текста).

Системы, использующие ранее заготовленные образцы, появились раньше других. Их суть проста, диктор заранее произносит фразы или предложения, посредством которых компьютер «обучается», и уже когда диктор произносит текст, сравнивает речь и заготовленные образцы и уже на основе этого сравнения соотносит речи и текст. Но за простотой кроется главный недостаток: система может точно интерпретировать только те команды, которые произнесены одним человеком, с одинаковой интонацией, скоростью и другими особенностями произношения.

С точки зрения лингвистики, наибольший интерес представляют системы, которые выделяют из речи лексические элементы. С помощью аппаратно-программного комплекса из речи выделяются фонемы и аллофоны, сопоставляя которые система получает наполненный смыслом текст. А вот лексическую составляющую этих систем мы разберем подробнее.

Распознавание слитной речи представляет собой многоуровневый процесс. Как всем известно, человеческая речь является звуком. Чтобы научить компьютер «понимать» речь, нужно звук преобразовать в электрический сигнал, который будет ему понятен. Колебания воздуха преобразуются в переменный ток в микрофоне. Затем АЦП производит дискретизацию сигнала. Но полученный сигнал нельзя сразу обрабатывать – нужно провести предварительную обработку, а именно: нормализовать (масштабирование амплитуды сигнала до масштаба текущей разрядности системы), обработать с помощью фильтров верхних, нижних частот, а также полосовым фильтром. Только после этого полученный сигнал можно анализировать на предмет выделения лексических составляющих. Это первый уровень распознавания. Затем в дело вступают нейронные сети, благодаря гибкости которых становится возможным распознавание отдельных речевых примитивов.

Далее следует второй уровень, на котором из результатов предыдущего выделяются слоги и морфемы, а следом за этим, на третьем этапе, слова и предложения.

При переходе на каждый следующий уровень обработки, кроме результатов классификации, добавляется информация о временных зависимостях и как сигналы соотносятся друг с другом. На высших уровнях происходит накопление данных с низших и их последующая обработка. Кроме того, высшие уровни могут управлять низкими, например с помощью «механизма внимания».

Для того, чтобы наглядно выделить из сигнала составные части речи необходимо воспользоваться осциллографом, с помощью которого мы сможем измерить параметры, такие как амплитуда, частота, длительность импульсов и период их следования и так далее. Затем в процессе изучения осциллограммы можно выделить аллофоны, как конкретную реализацию фонем. В системе распознания речи эти действия выполняются математическими способами.

На осциллограмме мы можем увидеть, как меняются вышеуказанные параметры со временем. С помощью спектрального анализа сигнала, мы получим информацию о каждой частоте и ее интенсивности. Благодаря этим данным, на спектрограмме можно увидеть отдельные примитивы речи. Но мало их обнаружить – гораздо сложнее их классифицировать. Ведь, кроме того, что фонема может иметь большое количество аллофонов, последние в свою очередь могут менять оттенок звучания в зависимости от внешних условий, наличия шума, и даже психологического состояния диктора.

Нейронная сеть самообучаема – ей не требуется учитель. В процессе обучения формируются так называемые нейронные ансамбли, которые образуются после статистической обработки всех поступающих сигналов. Т.е. каждый ансамбль соответствует наиболее часто встречающимся сигналам. Запоминание редких сигналов происходит позже и требует подключения механизма внимания или иного контроля высшего уровня.

После выделения информативных признаков речевого сигнала можно представить эти признаки в виде некоторого набора числовых параметров (т.е. в виде вектора в некотором числовом пространстве). Далее задача распознавания примитивов речи (фонем и аллофонов) сводится к их классификации при помощи обучаемой нейронной сети.

При этом обучение выделению примитивов речи (фонем и аллофонов) может заключаться в формировании нейронных ансамблей, ядра которых соответствуют наиболее частой форме каждого примитива.

Нейронные сети можно использовать и на более высоких уровнях распознавания слитной речи для выделения слогов, морфем и слов.

Нейронные сети сильно увеличили гибкость и точность систем, но в последнее время их развитие приостановилось. Казалось бы причина в недостаточной вычислительной мощности компьютеров. Но на деле все оказалось намного сложнее. Человеческий мозг может создавать речь, используя лишь правила функциональной грамматики и семантическую парадигму каждого слова в текущей ситуации. С помощью этих правил мозг понимает, какие слова могут сочетаться друг с другом, а какие нет. Человеку даже иногда не нужно слышать всю речь целиком – достаточно обрывков и знания предметной области.

Несмотря на то, что уровень точности распознавания у уже созданных систем достиг 80%, этого недостаточно для приближения к человеческим показателям, что составляют 96-98%. В настоящее время многие ученые, программисты, лингвисты согласны с тем, что существующие методы обработки не дают достаточной информации для распознавания речи, и они видят задачу в создании новой модели человеческой речи, в более углубленном изучении лингвистического аспекта. И когда это будет выполнено, распознание речи станет огромным скачком на пути создания искусственного интеллекта.